

**ANALISA SENTIMEN MENGGUNAKAN NAÏVE BAYES
UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP
KENAIKAN HARGA JUAL ROKOK PADA MEDIA SOSIAL
TWITTER**



**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada
Jurusan Informatika Fakultas Komunikasi dan Informatika**

Oleh:

FAUZIAH AFSHOH

L 200 130 034

**PROGRAM STUDI INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA
2017**

HALAMAN PERSETUJUAN

**ANALISA SENTIMEN MENGGUNAKAN NAÏVE BAYES
UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP
KENAIKAN HARGA JUAL ROKOK PADA MEDIA SOSIAL
TWITTER**

PUBLIKASI ILMIAH

oleh:

FAUZIAH AFSHOH

L 200 130 034

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



Endang Wahyu Pamungkas, S.Kom, M.Kom.

NIK. 100.1704

HALAMAN PENGESAHAN

**ANALISA SENTIMEN MENGGUNAKAN NAÏVE BAYES
UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP
KENAIKAN HARGA JUAL ROKOK PADA MEDIA SOSIAL
TWITTER**

OLEH


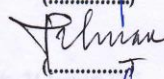

FAUZIAH AFSHOH

L 200 130 034

Telah dipertahankan di depan Dewan Penguji
Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta
Pada hari Selasa, 24 Januari 2017
Dan dinyatakan telah memenuhi syarat

Dewan Penguji:

1. Endang Wahyu P., S.Kom, M.Kom.
(Ketua Dewan Penguji)
2. Helman Muhammad, S.T., M.T.
(Anggota I Dewan Penguji)
3. Aris Rakhmadi, S.T., M.Eng.
(Anggota II Dewan Penguji)


(.....)

(.....)

(.....)

Mengetahui,


Dekan
Fakultas Komunikasi dan Informatika

Husni Thamrin, S.T., M.T., Ph.D.
NIK. 706


Ketua Program Studi
Informatika

Dr. Heru Supriyono, M.Sc.
NIK. 970

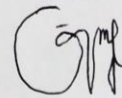
PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Naskah Publikasi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 24 Januari 2017

Penulis



FAUZIAH AFSHOH

L 200 130 034



UNIVERSITAS MUHAMMADIYAH SURAKARTA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: informatika@ums.ac.id

SURAT KETERANGAN LULUS PLAGIASI

012/A.3-IL.3/INF-FKI/I/2017

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : FAUZIAH AFSHOH
NIM : L200130034
Judul : ANALISA SENTIMEN MENGGUNAKAN NAÏVE BAYES
UNTUK MELIHAT PERSEPSI MASYARAKAT
TERHADAP KENAIKAN HARGA JUAL ROKOK PADA
MEDIA SOSIAL TWITTER
Program Studi : Informatika
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 02 Februari 2017

Biro Skripsi Informatika


Ihsan Cahyo Utomo, S.Kom., M.Kom.



UNIVERSITAS MUHAMMADIYAH SURAKARTA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: informatika@ums.ac.id

wisuda 2017 wisuda maret - DUE 17-Jan-2017 Roadmap Paper 14 of 38

Originality GradeMark PeerMark

ANALISA SENTIMEN MENGGUNAKAN NAÏVE BAYES UNTUK MELIHAT PERSEPSI
BY FAUZIAH AFSHOH

turnitin 14% SIMILAR OUT OF 0

ANALISA SENTIMEN MENGGUNAKAN NAÏVE BAYES UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP KENAikan HARGA JUAL ROKOK PADA MEDIA SOSIAL TWITTER

Abstrak

Media sosial *Twitter* merupakan salah satu contoh media sosial yang digunakan masyarakat untuk saling berinteraksi satu sama lain. *Twitter* memberikan layanan kepada penggunanya untuk mengirim dan membaca *tweets* yang telah dibagikan, sehingga masyarakat lebih memilih menuangkan opininya melalui media sosial daripada menyampaikannya secara langsung. Opini masyarakat yang tertuang dalam media sosial *twitter* berupa sebuah persepsi, baik itu positif maupun negatif. Melimpahnya opini masyarakat dapat dimanfaatkan sebagai bahan penelitian untuk mencari sebuah informasi. Pemanfaatan informasi tersebut membutuhkan teknik analisis yang tepat sehingga informasi yang dihasilkan mampu membantu banyak pihak dalam mengambil sebuah keputusan. Penggunaan teknik dalam pengolahan data dapat diselesaikan dengan teknik analisa sentimen atau *opinion mining*. Oleh sebab itu, pada penelitian ini mencoba melakukan analisa sentimen untuk melihat persepsi masyarakat terhadap kenaikan harga jual rokok pada media sosial *twitter* menggunakan metode Naive Bayes Classifier dengan mengklasifikasikan sentimen menjadi positif, negatif dan netral. Hasil penelitian yang telah dilakukan dapat diketahui bahwa sentimen positif paling banyak terbentuk dalam menanggapi wacana kenaikan harga jual rokok. Selain itu, hasil dari pengujian kinerja sistem menggunakan data latih 150 positif, 150

Match Overview

Rank	Source	Similarity
1	repository.widyatama.a...	2%
2	ojs.unud.ac.id	2%
3	id.scribd.com	1%
4	dergipark.ulakbim.gov.tr	1%
5	waset.org	1%
6	e-journal.uajy.ac.id	1%
7	fb.docs.com	1%
8	Submitted to Universit...	1%

PAGE: 1 OF 13 Text-Only Report

ANALISA SENTIMEN MENGGUNAKAN NAÏVE BAYES UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP KENAIKAN HARGA JUAL ROKOK PADA MEDIA SOSIAL TWITTER

Abstrak

Media sosial *Twitter* merupakan salah satu contoh media sosial yang digunakan masyarakat untuk saling berinteraksi satu sama lain. *Twitter* memberikan layanan kepada penggunanya untuk mengirim dan membaca *tweets* yang telah dibagikan, sehingga masyarakat lebih memilih menuangkan opininya melalui media sosial daripada menyampaikannya secara langsung. Opini masyarakat yang tertuang dalam media sosial *twitter* berupa sebuah persepsi, baik itu positif maupun negatif. Melimpahnya opini masyarakat dapat dimanfaatkan sebagai bahan penelitian untuk mencari sebuah informasi. Pemanfaatan informasi tersebut membutuhkan teknik analisis yang tepat sehingga informasi yang dihasilkan mampu membantu banyak pihak dalam mengambil sebuah keputusan. Penggunaan teknik dalam pengolahan data dapat diselesaikan dengan teknik analisa sentimen atau *opinion mining*. Oleh sebab itu, pada penelitian ini mencoba melakukan analisa sentimen untuk melihat persepsi masyarakat terhadap kenaikan harga jual rokok pada media sosial *twitter* menggunakan metode Naïve Bayes Classifier dengan mengklasifikasikan sentimen menjadi positif, negatif dan netral. Hasil penelitian yang telah dilakukan dapat diketahui bahwa sentimen positif paling banyak terbentuk dalam menanggapi wacana kenaikan harga jual rokok. Selain itu, hasil dari pengujian kinerja sistem menggunakan data latih 150 positif, 150 negatif dan 50 netral dengan metode Naïve Bayes Classifier menghasilkan nilai akurasi klasifikasi yang lebih baik daripada menggunakan metode Lexicon Based.

Kata Kunci: Analisa Sentimen, Naïve Bayes Classifier, Persepsi, Twitter.

Abstract

Social media Twitter is one example of social media that allows people to interact with each other. Twitter provides services to its users to send and read tweets that have been shared, so people prefer to pour of their opinions through social media rather than pass them directly. Public opinion contained in social media twitter be a perception, whether it is positive or negative. The huge amount of public opinion can be used as research material to locate information. Utilization of such information requires proper analysis techniques so that the resulting information can help some parties to take a decision. The use of the techniques in data processing can be completed using sentiment analysis or opinion mining. Therefore, in this study tries to analyze sentiment to see the public perception of the increase in cigarette prices on social media twitter using Naïve Bayes classifier to classify sentiment becomes positive, negative and neutral. The results of research that has been done can be seen that most of the positive sentiment was formed in response to the discourse of the increase in cigarette prices. In addition, results from testing the performance of the system using training data 150 positive, 150 negative and 50 neutral with Naïve Bayes classifier method produces a value classification accuracy better than using methods Lexicon Based.

Keywords: Sentiment Analysis, Naïve Bayes Classifier, Perception, Twitter.

1. PENDAHULUAN

Perkembangan akses teknologi informasi yang semakin pesat pada era globalisasi saat ini dapat memberikan kemudahan untuk berkomunikasi lebih efektif dan efisien, contohnya adalah *internet*. Seperti halnya yang telah diungkapkan oleh Kementrian Komunikasi dan Informatika (Kemenkominfo, 2013) bahwa pengguna *internet* di Indonesia saat ini mencapai 63 juta orang. Dari angka tersebut, terdapat 95 persennya menggunakan *internet* untuk mengakses media sosial. Salah satu media sosial yang diakses masyarakat adalah *Twitter*.

Media sosial *twitter* merupakan salah satu situs *microblogging* yang memungkinkan penggunanya untuk menulis tentang berbagai topik dan membahas isu-isu yang sedang terjadi. *Twitter* memberikan layanan kepada penggunanya untuk mengirim atau membaca *tweets* yang telah dibagikan dengan membatasi karakter maksimal 140. Adanya layanan tersebut menyebabkan masyarakat lebih memilih menuangkan opininya melalui media sosial daripada menyampaikannya secara langsung. Opini-opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi. Namun, dalam pemanfaatannya membutuhkan analisis yang tepat sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan.

Text mining adalah sebuah proses pengambilan informasi dari data tekstual yang memiliki kualitas tinggi serta dapat mengetahui permasalahan dalam teks dari sebuah topik tertentu. *Text mining* dalam analisa sentimen mampu mengidentifikasi emosional terhadap suatu pernyataan (Zhang, et al, 2015). Semenjak penelitian dari (Pang, Lee, 2008), banyak ditemukan studi pembelajaran tentang analisa sentimen. Analisa sentimen sering juga dikenal dengan *opinion mining* adalah studi komputasi dari pendapat, sentimen, sikap, serta emosi yang disajikan dalam sebuah teks (Liu, B., 2012). Seperti pada penelitian (Patela, T., Undavia, J., & Patela, A., 2015) melakukan analisa sentimen terhadap *feedback* yang diberikan oleh orang tua murid dengan membagi sentimen menjadi 3 kategori yaitu sentimen positif, negatif dan netral.

Salah satu teknik pembelajaran dari *text mining* untuk analisa sentimen adalah Naïve Bayes Classifier. Metode Naïve Bayes Classifier dianggap sebagai metode yang berpotensi baik untuk melakukan klasifikasi data daripada metode klasifikasi lainnya dalam hal akurasi dan komputasi (Ting, S. L., Ip, W. H., Tsang, A. H.C., 2011)(Joshi, M., & Vala, H., 2014). Algoritma Naïve Bayes Classifier dapat digunakan untuk memprediksi suatu nilai dari variabel dalam *data testing* (Nurrohmat, M. A., 2015). Oleh karena itu, penelitian ini mencoba melakukan analisa sentimen menggunakan Naïve Bayes untuk melihat persepsi masyarakat terhadap kenaikan harga jual rokok pada media sosial *twitter*. Tujuannya yaitu untuk memprediksi apakah masyarakat lebih banyak yang beranggapan positif, negatif, ataupun netral, sehingga dari hasil prediksi tersebut pemerintah dapat menentukan kebijakannya terhadap wacana kenaikan harga jual rokok. Selain itu, penelitian

dilakukan untuk mengetahui akurasi klasifikasi sistem menggunakan metode Naïve Bayes Classifier.

2. METODOLOGI



Diagram di atas merupakan proses metodologi penelitian yang akan dilakukan dalam penelitian. *Input* data yang diproses berupa kalimat opini yang diambil dengan melihat persepsi masyarakat tentang wacana kenaikan harga jual rokok melalui media sosial *twitter*. Proses selanjutnya yaitu melakukan data *preprocessing* yang terdiri dari *tokenizing*, *cleansing*, *normalization* dan *case folding*. Kalimat opini yang telah dilakukan proses *preprocessing* berupa kalimat dalam bentuk teks Bahasa Indonesia yang mudah dipahami. Setelah itu kalimat akan diproses dengan melakukan seleksi fitur *Chi Square*. Langkah selanjutnya akan dilakukan proses klasifikasi menggunakan algoritma Naïve Bayes Classifier.

Adapun penjelasan untuk serangkaian proses diagram di atas, yaitu:

1. *Input* data

Data *twitter* yang berisi kalimat opini dari masyarakat akan diambil secara acak dengan menggunakan metode pengumpulan *tweets* manual yaitu memasukkan berbagai macam kata kunci yang berhubungan dengan wacana kenaikan harga jual rokok pada kolom pencarian *twitter* dengan menggunakan *hashtag* seperti #rokokmahal yang diambil data sebanyak 120 buah yaitu 50 opini positif, 50 opini negatif dan 20 opini netral, #rokok sebanyak 100 buah yaitu 50 opini positif, 45 opini negatif, 5 opini netral, #KenaikanHargaRokok sebanyak 55 buah yaitu 30 opini positif, 25 opini negatif, #PolemikHargaRokok sebanyak 30 buah yaitu 10 opini positif, 10 opini negatif dan 10 opini netral, #ILCRokok50Ribu 20 buah yaitu 5 opini positif, 5 opini negatif dan 10 opini netral, serta data yang diambil berdasarkan *mention* yang masuk pada pengguna *twitter* @jokowi dan @ILC sebanyak 25 buah yaitu 5 opini positif, 15 opini negatif dan 5 opini netral. Total data *twitter* yang diperoleh yaitu sebanyak 350 buah.

2. Data *preprocessing*

Tahap *preprocessing* merupakan tahap awal untuk mengolah data teks menjadi analisa sentimen. Terdapat beberapa metode yang digunakan pada tahap *preprocessing* seperti :

1. *Parsing / Tokenizer* merupakan proses untuk memisahkan data teks menjadi beberapa token.
 2. *Cleansing* adalah proses pembersihan data teks yaitu dengan menghilangkan data yang tidak konsisten atau tidak relevan.
 3. *Normalization* digunakan untuk menormalisasikan kalimat. Banyak pengguna dalam menuliskan kata menggunakan singkatan sehingga akan menghambat proses pemahaman teks, seperti “BGT” jika dinormalisasi menjadi “BANGET”, kata perulangan “gara2” menjadi “gara-gara”.
 4. *Case folding* yaitu merubah bentuk kata menjadi bentuk dasarnya agar sebuah karakter dapat seragam (*lower case*).
3. Seleksi fitur *Chi Square*

Proses seleksi fitur *Chi Square* dilakukan untuk membersihkan fitur-fitur yang mengganggu dalam tahap klasifikasi menggunakan metode Naïve Bayes Classifier. Seleksi fitur *Chi Square* yaitu suatu konsep pembelajaran ilmu statistika yang dapat digunakan untuk mengamati kesesuaian dari *terms* dengan kategorinya. Untuk melakukan perhitungan nilai *Chi Square* dapat menggunakan rumus berikut.

$$X^2(D, t, c) = \frac{(N_{00} + N_{11} + N_{10} + N_{01})x(N_{00}N_{11} - N_{10}N_{01})^2}{(N_{11} + N_{01})x(N_{11} + N_{10})x(N_{10} + N_{00})x(N_{01} + N_{00})}$$

Dimana, nilai:

N_{00} = jumlah data yang tidak termasuk kategori c dan tidak mengandung t

N_{10} = jumlah data yang termasuk kategori c dan tidak mengandung t

N_{01} = jumlah data yang tidak termasuk kategori c dan mengandung t

N_{11} = jumlah data yang termasuk kategori c dan mengandung t

4. Analisa Sentimen dengan Naïve Bayes Classifier

Naïve Bayes Classifier merupakan teknik pembelajaran algoritma data mining yang memanfaatkan metode probabilitas dan statistik. Naïve Bayes Classifier dalam melakukan klasifikasi terdapat dua proses penting yaitu *learning (training)* dan *testing*. Penelitian ini menggunakan *data training* dari opini masyarakat yang diperoleh yaitu sebanyak 350 buah dengan kategori opini positif 150 buah, opini negatif 150 buah dan opini netral 50 buah beserta klasifikasinya secara manual atau data yang sudah diketahui kategorinya. Sedangkan *data testing* digunakan untuk prediksi klasifikasi atau data yang belum diketahui klasifikasinya.

Proses klasifikasi Naïve Bayes Classifier terhadap dokumen yaitu dengan mempresentasikan setiap dokumen dengan atribut “ $X_1, X_2, X_3, \dots, X_n$ ” yang mempunyai makna bahwa X_1 untuk kata pertama, X_2 adalah kata kedua, dan seterusnya. Untuk himpunan kategori *tweet* dipresentasikan dengan V . Saat melakukan proses klasifikasi dokumen, Naïve Bayes Classifier akan mencari nilai probabilitas tertinggi dari:

$$V_{MAP} = \underset{V_j \in V}{argmax} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \quad (1)$$

Jika nilai dari $P(X_1, \dots, X_n)$ adalah konstan untuk semua kategori V_j maka dari persamaan (1) dapat ditulis :

$$V_{MAP} = \underset{V_j \in V}{argmax} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \quad (2)$$

Sehingga dari persamaan (2) dapat ditulis sebagai :

$$V_{MAP} = \underset{V_j \in V}{argmax} \prod_{i=1}^n P(x_i | V_j) P(V_j) \quad (3)$$

Keterangan :

V_{MAP} : semua kategori yang diujikan

V_j : kategori tweet, dengan :

j_1 : sentimen positif

j_2 : sentimen negatif

j_3 : sentimen netral

$P(X_i / V_j)$: probabilitas X_i pada kategori V_j

$P(V_j)$: probabilitas dari V_j

Pada saat pelatihan untuk $P(X_i / V_j)$ menggunakan rumus *Chi Square* yaitu :

$$P(X_i | V_j) = \frac{(N_{00} + N_{11} + N_{10} + N_{01})x(N_{00}N_{11} - N_{10}N_{01})^2}{(N_{11} + N_{01})x(N_{11} + N_{10})x(N_{10} + N_{00})x(N_{01} + N_{00})}$$

Dan nilai $P(V_j)$ dapat dihitung menggunakan persamaan berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh|}$$

Keterangan:

$|docs\ j|$ = jumlah dokumen setiap kategori j

$|contoh|$ = jumlah dokumen dari semua kategori

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Sistem analisa sentimen menggunakan data opini masyarakat berbahasa Indonesia berdasarkan topik kenaikan harga jual rokok pada media sosial *twitter*, memiliki spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk mendukung proses penelitian yaitu:

1. Satu buah laptop yang memiliki spesifikasi sebagai berikut :
 - Processor AMD A8-6410 APU with AMD Radeon R5 Grapichs 2.00 GHz, RAM 4 GB, Hardisk 500 GB, DVDRW, Wi-Fi, OS Windows 8.1.
2. *Software* yang digunakan adalah :
 - NetBeans IDE 8.0.2, *Java*.

Hasil analisis dan pengujian yang telah dilakukan menggunakan data latih sebanyak 350 buah dari opini masyarakat mengenai topik kenaikan harga jual rokok menggunakan metode Lexicon Based menghasilkan nilai *precision* 53% pada data opini negatif dan *recall* sebesar 57% pada data opini negatif, kemudian untuk nilai *accuracy* menggunakan Lexicon Based menghasilkan nilai persentase sebesar 81% pada data opini netral. Metode Lexicon Based mampu mengatasi dokumen yang sama tetapi memiliki sentimen berbeda dalam konteks yang berbeda (Pamungkas, E. W., & Putri, D. P., 2016). Selain itu metode Lexicon Based dalam melakukan pengujiannya dapat memberikan nilai *precision* yang tinggi tetapi memberikan nilai *recall* yang rendah (Khan, A. Z., Atique, M., & Tharake, V. M., 2015). Hal ini terbukti ketika metode Lexicon Based mencari nilai *precision* dan *recall* pada tipe data opini netral. Sedangkan pengujian menggunakan metode Naïve Bayes Classifier, dapat memberikan nilai persentase *precision*, *recall* dan *accuracy* yang lebih tinggi daripada Lexicon Based, karena metode Naïve Bayes Classifier dalam melakukan pengujian klasifikasi dokumen dapat memiliki tingkat ketepatan klasifikasi yang tinggi meskipun sangat sederhana dan efisien yaitu hanya dengan memanfaatkan penggunaan kemunculan dan frekuensi fitur pada setiap opini. Selain itu, hasil dari ketepatan klasifikasi menggunakan metode Naïve Bayes Classifier bergantung pada jumlah data latih yang digunakan dalam sistem. Penggunaan data latih dalam pengujian yaitu 150 buah positif, 150 buah negatif dan 50 netral dengan melakukan teknik pengujian sebanyak sepuluh kali atau *Ten Fold Cross Validation* (10-fold) yaitu dengan membagi data menjadi 10 bagian. Data yang diujikan masing-masing berisi 15 data positif, 15 data negatif dan 5 data netral yang diambil secara acak atau setiap bagian berisi data yang berbeda dari data sebelumnya. Hasil analisis dari pengujian data menggunakan metode Naïve

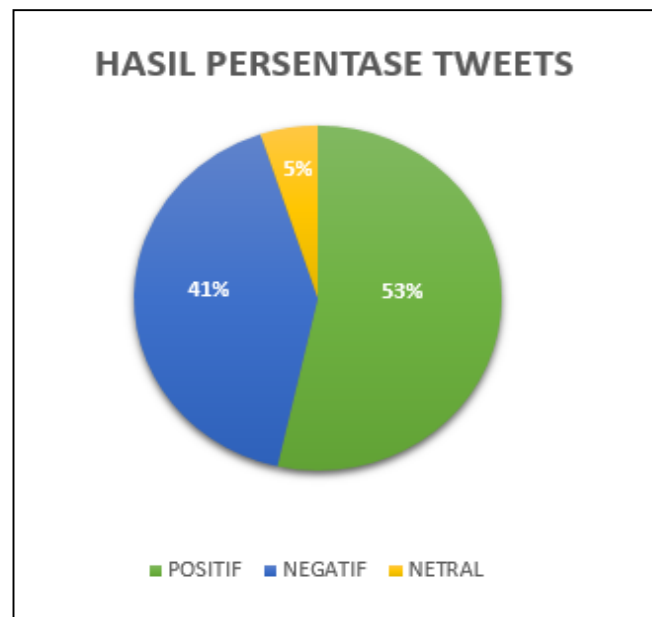
Bayes Classifier, diperoleh nilai *precision* terbesar yaitu 76% pada data uji ke 76 sampai 90 dengan tipe opini positif. Nilai terbesar dari *recall* yaitu 93% dengan tipe opini positif pada pengujian data uji ke 136 sampai 150. Nilai terbesar dari *accuracy* pada saat melakukan pengujian yaitu 88% dengan opini netral pada data uji ke 91 sampai 105 dan data uji 121 sampai 135. Setelah melakukan pengujian data pada setiap bagian, berikutnya menghitung nilai rata-rata dari nilai *precision*, *recall* dan *accuracy*. Berdasarkan analisa diatas, dapat disimpulkan bahwa kinerja Naïve Bayes Classifier dalam melakukan klasifikasi dokumen terbukti lebih baik daripada menggunakan metode klasifikasi Lexicon Based. Hasil tersebut dapat dilihat pada Tabel 1. (Kusumawati, I., 2017).

Tabel 1. Hasil pengujian

Opinion Type	Classification					
	Naïve Bayes Classifier			Lexicon Based		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
Positive	65%	80%	72%	51%	56%	58%
Negative	62%	64%	67%	53%	57%	60%
Neutral	34%	17%	85%	10%	4%	81%

Pada Tabel 1 dapat dilihat bahwa nilai *precision* data opini positif menggunakan metode Naïve Bayes Classifier menunjukkan nilai persentase yang lebih besar daripada nilai *precision* menggunakan metode Lexicon Based yaitu 65%, hal ini disebabkan nilai tingkat ketepatan data opini positif antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem menunjukkan nilai *true positive* atau *correct result* yang lebih besar daripada nilai opini yang lainnya. Nilai persentase terbesar dari *recall* yaitu 80% pada tipe opini positif dengan metode klasifikasi yang digunakan yaitu Naïve Bayes Classifier, besarnya nilai persentase *recall* disebabkan oleh tingkat keberhasilan sistem pada data opini positif dalam menemukan kembali sebuah informasi yaitu hanya memiliki sedikit kesalahan saat proses klasifikasi atau *missing result*. Sedangkan nilai persentase *accuracy* terbesar berdasarkan tabel diatas yaitu 85% pada tipe opini netral dengan metode klasifikasi Naïve Bayes Classifier. Data yang tidak seimbang antara opini netral dengan opini positif dan negatif inilah yang mengakibatkan nilai *accuracy* meningkat, karena data yang diluar netral jumlahnya lebih banyak daripada data yang diluar positif dan negatif atau nilai dari *true*

negative atau *correct absence of result* lebih besar dari opini positif dan negatif. Hasil penelitian yang telah dilakukan dapat diketahui bahwa nilai sentimen yang paling banyak terbentuk dalam menanggapi wacana kenaikan harga jual rokok adalah sentimen positif. Kategori sentimen positif merupakan kategori opini masyarakat yang setuju dengan adanya wacana kenaikan harga jual rokok, sehingga dari hasil inilah pemerintah dapat mengambil kebijakannya. Nilai persentase dari sentimen positif dan negatif memiliki selisih angka yang tidak cukup jauh, karena data latih dari sentimen positif dan negatif jumlahnya sama yaitu data positif 150 buah dan data negatif 150 buah.



Gambar 1. Hasil persentase *tweets*

3.2 Pembahasan

Metode penelitian yang dimulai dari pengumpulan *tweets* sampai melakukan perhitungan nilai probabilitas menggunakan Naïve Bayes Classifier masih terdapat berbagai permasalahan yang membuat penelitian ini bekerja tidak maksimal, yaitu:

1. Ketidakseimbangan Data

Data yang tidak seimbang mengakibatkan hasil dari klasifikasi sistem tidak memuaskan. Hal ini karena data *twitter* yang digunakan masing-masing berjumlah 150 positif, 150 negatif dan 50 netral. Jumlah data *twitter* netral tidak seimbang dengan data *twitter* positif dan negatif sehingga pada saat melakukan klasifikasi sistem banyak ditemukan dari data *twitter* netral cenderung terklasifikasi sebagai data positif ataupun negatif. Misalnya pada

kalimat “Sedikit saran untuk bapak dan ibu pemerintah seharusnya memikirkan hal yang lebih penting daripada menaikkan harga rokok”. Kalimat tersebut sebenarnya terklasifikasi menjadi kalimat netral, namun setelah dilakukan pengujian menggunakan sistem ternyata kalimat tersebut masuk dalam klasifikasi kalimat negatif. Adanya permasalahan tersebut yang mengakibatkan penelitian ini bekerja kurang maksimal.

2. Terdapat Dataset yang Tidak Tepat

Terdapat kesalahan dalam proses klasifikasi sistem yang disebabkan oleh penggunaan dataset yang kurang tepat. Keadaan ini yang menyebabkan banyak ditemukan fitur kemunculan data yang bukan termasuk kategorinya pada data uji yang digunakan. Contohnya dataset yang digunakan dalam data positif terdapat kata “setuju”, kemudian pada data negatif juga terdapat kata “setuju” yang berasal dari kata “tidak setuju”. Kata yang sama pada dataset positif dan negatif inilah yang mengakibatkan kesalahan dalam proses klasifikasi data. Kalimat “Saya tidak setuju harga rokok naik” yang sebenarnya terklasifikasi menjadi kalimat negatif, tetapi pada saat sistem bekerja kalimat tersebut akan masuk dalam klasifikasi positif, karena dalam kalimat tersebut terdapat kata “setuju” dan kata setuju ada dalam dataset positif. Jadi dapat disimpulkan bahwa permasalahan penggunaan dataset yang didalamnya terdapat kata yang sama akan mempengaruhi proses klasifikasi pada sistem. Permasalahan tersebut juga dapat memicu tingkat akurasi sistem, sehingga mengakibatkan kinerja sistem tidak berjalan secara maksimal.

3. Kendala Memahami Klasifikasi Sentimen

Kendala yang sering kali ditemui dalam memahami kalimat untuk diolah dalam proses klasifikasi sentimen yaitu ketika menjumpai kalimat awal yang memberikan persepsi dalam kategori positif atau negatif kemudian diakhir kalimat memberikan persepsi bahwa kalimat tersebut termasuk dalam kategori netral, ataupun kalimat yang awalnya memberikan persepsi netral kemudian diakhir kalimat memberikan persepsi positif ataupun negatif. Kendala tersebut yang mengakibatkan kinerja sistem tidak berjalan secara maksimal karena sistem akan mendeteksi klasifikasi sentimen pada kalimat yang berada di awal saja. Salah satu contohnya yaitu pada kalimat *data training* “Orang biasa meninggal itu bukan karena rokok, tetapi orang yang meninggal itu sudah diatur oleh Allah. Saya tidak setuju harga rokok naik”. Dalam klasifikasi data manual kalimat tersebut masuk dalam kategori sentimen negatif tetapi pada saat klasifikasi menggunakan

sistem, kalimat tersebut masuk dalam kategori sentimen positif karena pada awal kalimat terdapat kata yang mengandung sentimen positif.

4. PENUTUP

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa:

1. Data yang tidak seimbang mengakibatkan proses klasifikasi tidak berjalan secara maksimal, untuk penelitian lebih lanjut dapat menggunakan data yang seimbang agar klasifikasi dokumen mampu bekerja dengan baik.
2. Permasalahan dalam penggunaan dataset yang tidak tepat dapat menggunakan penggabungan metode Naïve Bayes Classifier dengan berbagai fitur n-gram sehingga ketepatan hasil akurasi klasifikasi akan lebih baik.
3. Proses klasifikasi dokumen yang memiliki kendala dalam memahami klasifikasi sentimen dapat diselesaikan dengan metode *negation detection*.
4. Penelitian yang telah dilakukan menggunakan data opini sebanyak 350 buah dapat diketahui bahwa nilai sentimen positif yang paling banyak terbentuk dalam menanggapi topik kenaikan harga jual rokok dengan persentase *tweets* sebesar 53%. Sedangkan untuk evaluasi kinerja sistem menggunakan Naïve Bayes Classifier menunjukkan hasil klasifikasi dokumen yang lebih baik daripada metode Lexicon Based.

DAFTAR PUSTAKA

- Joshi, M., & Vala, H. (2014). Opinion Mining For Sentiment Data Classification. *International Journal of Research in Information Technology*, 3(1), 1-13.
- Khan, A. Z., Atique, M., & Thakare, V. M. (2015). Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCE)*, 89.
- Ling, J., N Kencana, I. P. E., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92-99
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures On Human Language Technologies*, 5(1), 1-167.
- Nurrohmat, M. A. (2015). Aplikasi Pemrediksi Masa Studi dan Predikat Kelulusan Mahasiswa Informatika Universitas Muhammadiyah Surakarta Menggunakan Metode Naïve Bayes. *Khazanah Informatika*, (1).

- Pamungkas, E. W., & Putri, D. P. (2016). *An Experimental Study of Lexicon-Based Sentiment Analysis on Bahasa Indonesia*. Proceeding of The 6th International Annual Engineering Seminar (INAES).
- Pan, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Patel, T., Undavia, J., & Patela, A. (2015). Sentiment Analysis of Parents Feedback for Educational Institutes. *International Journal of Innovative and Emerging Research in Engineering*, 2(3).
- Ting, S. L., Ip, W. H., & Tsang, A. H. (2011). Is Naïve Bayes a Good Classifier for Document Classification?. *International Journal of Software Engineering and Its Applications*, 5(3), 37-46.